

PATENT ABSTRACTS OF JAPAN

(11) Publication number : 11-175523
(43) Date of publication of application : 02.07.1999

(51) Int. Cl. G06F 17/27
G06F 17/28

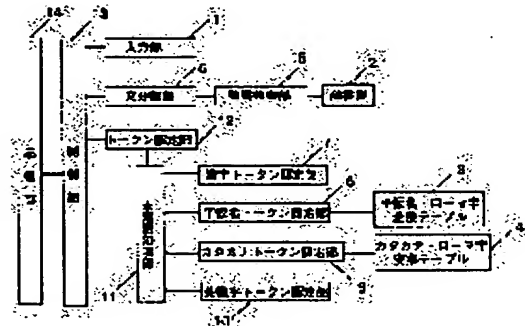
(21) Application number : 09-338400 (71) Applicant : MATSUSHITA ELECTRIC IND CO LTD
(22) Date of filing : 09.12.1997 (72) Inventor : KINOSHITA HITOMI

(54) DEVICE FOR ANALYZING JAPANESE MORPHEME AND METHOD THEREFOR AND RECORDING MEDIUM FOR STORING JAPANESE MORPHEME ANALYZING PROGRAM

(57) Abstract:

PROBLEM TO BE SOLVED: To correctly analyze a word which is not registered in a dictionary in an input sentence in the morpheme analysis of Japanese.

SOLUTION: This analyzing device is provided with an inputting part 1 for storing an inputted Japanese character string, dictionary retrieving part 5 for retrieving a dictionary group 2 which stores vocabulary information necessary for morphemic analysis by using the Japanese character string as a key, sentence dividing part 6 for dividing an input sentence stored in the inputting part 1 into morphemes by using the dictionary retrieving part 5, unknown word processing part 11 for judging whether or not a word is formed by linking the morpheme with another morpheme when the morpheme which is formed of one character is present for the morphemes divided by the sentence dividing part 6, and token recognizing part 12 for recognizing each divided unit as a morpheme based on the result of the unknown word processing part 11.



LEGAL STATUS

[Date of request for examination]
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

Copyright (C) ; 1998, 2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-175523

(43) 公開日 平成11年(1999) 7月2日

(51) Int.Cl.⁸

識別記号

F I

G 0 6 F 17/27
17/28

G 0 6 F 15/38

E
C

審査請求 未請求 請求項の数 6 O L (全 10 頁)

(21) 出願番号 特願平9-338400

(22) 出願日 平成9年(1997)12月9日

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 木下 ひとみ

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

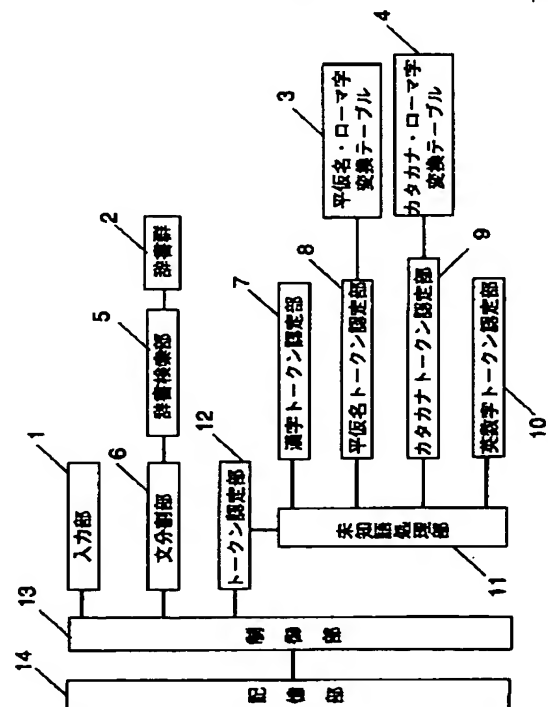
(74) 代理人 弁理士 滝本 智之 (外1名)

(54) 【発明の名称】 日本語形態素解析装置および日本語形態素解析方法ならびに日本語形態素解析プログラムを記録した記録媒体

(57) 【要約】

【課題】 日本語の形態素解析において、入力文中における辞書に未登録の単語を正しく解析すること。

【解決手段】 入力される日本語文字列を記憶する入力部1と、形態素解析に必要な語彙情報を記憶した辞書群2を用い、日本語文字列をキーとして辞書群2を検索する辞書検索部5と、辞書検索部5を用いて入力部1に記憶された入力文を形態素に分割する文分割部6と、文分割部6により分割された形態素に対し、1文字で形態素となるものが存在していれば、当該形態素と他の形態素とを連結して単語が形成されるのか否かを判断する未知語処理部11と、未知語処理部11の結果に基づいて分割された個々の単位を形態素として認定するトークン認定部12と、を有する。



【特許請求の範囲】

【請求項 1】入力される日本語文字列を記憶する入力文記憶手段と、

形態素解析に必要な語彙情報を記憶した辞書を用い、日本語文字列をキーとして辞書を検索する辞書検索手段と、

前記辞書検索手段を用いて、前記入力文記憶手段に記憶された入力文を形態素に分割する文分割手段と、

前記文分割手段により分割された形態素に対し、1文字で形態素となるものが存在していれば、当該形態素と他の形態素とを連結して単語が形成されるのか否かを判断する未知語処理手段と、

前記未知語処理手段の結果に基づいて分割された個々の単位を形態素として認定する形態素認定手段と、を有することを特徴とする日本語形態素解析装置。

【請求項 2】辞書登録見出しの読み情報をローマ字表記で記述しておくことにより、前記未知語処理手段で連結した形態素をローマ字表記で出力することを特徴とする請求項 1 記載の日本語形態素解析装置。

【請求項 3】平仮名とローマ字との対応関係を記憶した平仮名・ローマ字対応テーブルを有し、前記未知語処理手段は、1文字で形態素をなすものが平仮名である場合、前記平仮名・ローマ字対応テーブルを参照して、連結した平仮名をローマ字表記に変換することを特徴とする請求項 1 記載の日本語形態素解析装置。

【請求項 4】カタカナとローマ字との対応関係を記憶したカタカナ・ローマ字対応テーブルを有し、前記未知語処理手段は、1文字で形態素をなすものがカタカナである場合、前記カタカナ・ローマ字対応テーブルを参照して、連結したカタカナをローマ字表記に変換することを特徴とする請求項 1 記載の日本語形態素解析装置。

【請求項 5】日本語文を文字列として入力するステップと、

形態素解析に必要な語彙情報を記憶した辞書を参照しながら日本語文字列をキーとして辞書を検索し、入力文を形態素に分割するステップと、

前記ステップにおいて分割された形態素に対し、1文字で形態素となるものが存在していれば、当該形態素と他の形態素とを連結して単語が形成されるのか否かを判断するステップと、

前記ステップの判断結果に基づいて分割された個々の単位を形態素として認定する個々の単位をトークンとして認定するステップと、を有することを特徴とする日本語形態素解析方法。

【請求項 6】日本語文を文字列として入力するステップと、

形態素解析に必要な語彙情報を記憶した辞書を参照しながら日本語文字列をキーとして辞書を検索し、入力文を形態素に分割するステップと、

前記ステップにおいて分割された形態素に対し、1文字

で形態素となるものが存在していれば、当該形態素と他の形態素とを連結して単語が形成されるのか否かを判断するステップと、

前記ステップの判断結果に基づいて分割された個々の単位を形態素として認定する個々の単位をトークンとして認定するステップと、を有することを特徴とする日本語形態素解析プログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、文字列として入力された日本語文の形態情報を出力する日本語形態素解析装置および日本語形態素解析方法ならびに日本語形態素解析プログラムを記録した記録媒体に関するものである。

【0002】

【従来の技術】かな漢字変換や機械翻訳などにおいて日本語を処理する場合、まず、形態素解析を行う必要がある。形態素解析では、普通、単語をキーとしてその語彙情報を記憶した辞書を検索しながら、文字列を形態素（意味を持つ最小の単位であり、以下「トークン」と称する）に分割し（トークン分割）、個々のトークンに形態情報（品詞、活用など）を付加する。

【0003】日本語の形態素解析における辞書検索では、一般に 2 通りの手法が知られている。1 つは、入力文の任意の位置から始まる最も長い文字列（トークン）を獲得する方法で、最長一致の考え方である。しかしながら、最長一致法だけでは誤解釈を招くため、別の手法として、入力文の任意の位置から始まる全ての可能なトークンを獲得し、接続テーブルやトークンの優先度など、その他様々な情報を用いて、最適なトークン認定を行うものがより一般化している。

【0004】

【発明が解決しようとする課題】しかしながら、上記どちらの手法を用いても、辞書未登録語（以下「未知語」と称する）をどのように処理するかが大きな課題となる。例えば、人名や社名などの固有名詞を全て辞書に登録するのは不可能である。「田中」や「中村」のように、比較的好くある名であれば辞書に登録されているであろう。しかし、「上沼」や「古手川」のような珍しい名は、辞書に登録されていないことが多く、そのまま（「上沼」、「古手川」）では、未知語となる。そのような場合でも、通常、形態素解析で用いる辞書には、漢字 1 文字で辞書に登録されていることがあり、この 2 つの例の場合も、漢字 1 文字でトークンとなってしまう。この形態素解析結果を日英翻訳に用いると、「上沼」は「上の沼」と解釈され、「an uppermarsh」と訳出される。また、「古手川」は「古い手（川）」と解釈され、「old hand river」と訳出される。

【0005】本発明は、日本語の形態素解析において、入力文中における辞書に未登録の単語を正しく解析する

10

20

30

40

50

ことのできる日本語形態素解析装置および日本語形態素解析方法ならびに日本語形態素解析プログラムを記録した記録媒体を提供することを目的とする。

【0006】

【課題を解決するための手段】この課題を解決するために本発明は、入力される日本語文字列を記憶する入力文記憶手段と、形態素解析に必要な語彙情報を記憶した辞書を用い、日本語文字列をキーとして辞書を検索する辞書検索手段と、辞書検索手段を用いて、入力文記憶手段に記憶された入力文を形態素に分割する文分割手段と、文分割手段により分割された形態素に対し、1文字で形態素となるものが存在していれば、当該形態素と他の形態素とを連結して単語が形成されるのか否かを判断する未知語処理手段と、未知語処理手段の結果に基づいて分割された個々の単位を形態素として認定する形態素認定手段と、を有する構成とした。

【0007】これにより、日本語の形態素解析において、入力文中における辞書に未登録の単語について正しく形態素解析することのできる日本語形態素解析装置が得られる。

【0008】

【発明の実施の形態】本発明の請求項1に記載の発明は、入力される日本語文字列を記憶する入力文記憶手段と、形態素解析に必要な語彙情報を記憶した辞書を用い、日本語文字列をキーとして辞書を検索する辞書検索手段と、辞書検索手段を用いて、入力文記憶手段に記憶された入力文を形態素に分割する文分割手段と、文分割手段により分割された形態素に対し、1文字で形態素となるものが存在していれば、当該形態素と他の形態素とを連結して単語が形成されるのか否かを判断する未知語処理手段と、未知語処理手段の結果に基づいて分割された個々の単位を形態素として認定する形態素認定手段と、を有する構成としたことにより、入力文を分割した際に1文字毎に分割されるのを防ぐことができる。

【0009】本発明の請求項2に記載の発明は、請求項1に記載の発明において、辞書登録見出しの読み情報をローマ字表記で記述しておくことにより、未知語処理手段で連結した形態素をローマ字表記で出力する構成としたことにより、連結した単語についてローマ字表記で出力することができる。

【0010】以下、本発明の一実施の形態について、図面を参照しながら説明する。

（実施の形態）図1は本発明の一実施の形態における日本語形態素解析装置の機能ブロック図であり、機能手段による構成を示したものである。

【0011】図1において、1は操作者により日本語文字列を入力する入力部であり、本実施の形態では、入力される日本語文字列は漢字仮名混じり文としている。2は日本語文字列をキーとして、読み情報、品詞、活用の種類など、形態素解析に必要な語彙情報を記憶した辞書

群である。この辞書における読み情報は、平仮名で表記したもの（以下「平仮名読み」と称する）と、アルファベットで表記したもの（以下「ローマ字読み」と称する）との2つの情報を持つものとする。例えば、「上」という辞書見出しに関しては、平仮名読みとしては「うえ」、「うわ」、「かみ」および「じょう」を有し、ローマ字読みとしては、「ue」、「uwa」、「kami」および「jou」を有している。

【0012】3は平仮名1文字に対するローマ字表記を対応づけて記憶した平仮名・ローマ字変換テーブルである。4はカタカナ1文字に対するローマ字表記を対応づけて記憶したカタカナ・ローマ字変換テーブルである。本実施の形態では、平仮名・ローマ字変換テーブル3およびカタカナ・ローマ字変換テーブル4を、図3に示すような1つのテーブルとして対応している。

【0013】5は日本語文字列をキーとして、辞書群2を検索する辞書検索部である。6は辞書検索部5の検索結果を用いて、入力文を辞書登録単位に分割する文分割部である。

【0014】7は文分割部6で漢字1文字だけで分割される文字が存在する場合、前後の漢字文字と連結して固有名詞とするか否かを判断する漢字トークン認定部である。ここで、固有名詞と決定した文字列に対しては、個々の漢字のローマ字読みを結合して対訳文字列を生成する。

【0015】8は文分割部6で平仮名1文字だけで分割される文字が存在する場合、前後の平仮名文字と連結して固有名詞とするか否かを判断する平仮名トークン認定部である。ここで、固有名詞と決定した文字列に対しては、平仮名・ローマ字変換テーブル3を参照し、ここのローマ字表記を結合して、対訳文字列を生成する。

【0016】9は文分割部6でカタカナ1文字だけで分割される文字が存在する場合、前後のカタカナ文字と連結して固有名詞とするか否かを判断するカタカナトークン認定部である。ここで、固有名詞と決定した文字列に対しては、カタカナ・ローマ字変換テーブル4を参照し、ここのローマ字表記を結合して、対訳文字列を生成する。

【0017】10は文分割部6で英数字1文字だけで分割される文字が存在する場合、前後の英数文字と連結して固有名詞とするか否かを判断する英数字トークン認定部である。

【0018】11は漢字トークン認定部7、平仮名トークン認定部8、カタカナトークン認定部9、英数字トークン認定部10を制御する未知語処理部である。12は未知語処理部11を用いて、1文字で分割された文字を固有名詞として認定するトークン認定部である。ここでは、「再分割」のように、固有名詞ではないが、1つのまとまりと解釈できるトークンをまとめ上げる処理も行う。

【0019】13は入力部1、文分割部6、トークン認定部12を制御し、入力部1で入力された文の形態情報を出力する制御部である。14は入力部1で入力された文、文分割部6における分割結果、トークン認定部12における認定結果を記憶する記憶部である。

【0020】図2は本発明の一実施の形態における日本語形態素解析装置の回路ブロック図であり、ハードウェアによる構成を示している。

【0021】図2において、21はキーボードやポインティング・デバイスなどの入力装置であり、22は陰極線管ディスプレイ（CRT）や液晶ディスプレイ（LCD）などの表示装置である。

【0022】23は装置を制御する中央処理装置（CPU）、24はデータを一時的に記憶するランダム・アクセス・メモリ（RAM）、25はCPU23が実行するプログラムを格納するリー・ドオンリー・メモリ（ROM）である。

【0023】ここで、図1の機能手段と、図2のハードウェアとの対応関係を説明する。図1および図2において、入力部1は入力装置21により、記憶部14はRAM24によりそれぞれ実現されている。また、平仮名・ローマ字変換テーブル3およびカタカナ・ローマ字変換テーブル4は、ROM25に記憶され、辞書群2は、RAM24、ROM25および2次記憶装置26のいずれかに記憶されている。

【0024】辞書検索部5、文分割部6、漢字トークン認定部7、平仮名トークン認定部8、カタカナトークン認定部9、英数字トークン認定部10、未知語処理部11、トークン認定部12および制御部13は、CPU23がRAM24およびROM25とデータのやりとりを行いながら、ROM25に記憶された各種のプログラムを実行することにより実現されている。

【0025】なお、本実施の形態では、CPU23がROM25に記憶されるプログラムを実行する形態を示しているが、CPU23が実行するプログラムは、読取装置を用い、CD-ROM（コンパクト・ディスク・リード・オンリー・メモリ）などの記録媒体に記録されたプログラムを読み出して実行する形態であっても構わない。このように構成することにより、本発明を汎用コンピュータなどにおいて用意に実現が可能となる。

【0026】以上のように構成された日本語形態素解析装置について、以下にその動作を図4、図5および図7のフローチャートに基づいて説明する。なお、これらのフローチャートは、CPU23がROM25に記憶されたプログラムを実行する様子を示したものである。

【0027】図4、図5および図7は本発明の一実施の形態における日本語形態素解析装置の動作フローチャートであり、入力文を辞書群2を用いて辞書登録単位に分割し、辞書に未登録であった固有名詞を認定する形態素解析処理の流れを示したものである。

【0028】図4に示すように、まず、ステップS1では、入力部1より日本語文字列が入力される。ここで入力される文字列は、漢字仮名混じり文であるものとし、次のような文が入力されたものとする。

【0029】（a）私は、昨日、上沼さんに会いました。

ステップS2では、ステップS1において日本語文字列の入力があったかの否かをチェックし、入力があったのであればステップS3に進み、入力がなかったのであればステップS1に戻り、入力を待つ。

【0030】ステップS3では、ステップS1における入力文（a）を、辞書群2を用いて辞書登録単位に分割する。この処理の流れを図5に示すフローチャートに基づいて説明する。ここでは、最長一致法による解析法を用いる。また、辞書検索で得られた情報は、トークン・リストとして出力する。トークン・リストは、入力文（a）を辞書登録単位に分割した結果を双方向リストとして構造化したものである。

【0031】図5に示すように、まず、ステップT1では、変数“pos”、定数“len”および変数“S[n]”の初期化を行う。“pos”は、辞書検索キーの先頭文字を指すオフセットで、初期値は1である。“len”は、入力文の長さを表し、入力文（a）の場合の値は17である。“S[n]”は、入力文を指し、入力文のn文字目を“S[n]”と表現する（入力文（a）では、1文字目「私」はS[1]、4文字目「昨」はS[4]）。

【0032】ステップT2では、変数“pos”の値と定数“len”値とを比較し、“pos”が“len”より小さければステップT3に進み、そうでなければ図5に示す辞書検索による文分割処理を終える。

【0033】ステップT3では、変数“keyLen”に検索キーの長さを格納する。最初の検索キーの長さは、“len-pos+1”で与えられる。

【0034】ステップT4では、ステップT3で求めた“keyLen”が0（零）より大きいのか否かをチェックし、大きければステップT6に進み、そうでなければステップT5に進む。

【0035】ステップT5では、“S[pos]”を未知語として、トークン・リストに登録し、“pos”の値を1文字分進め、ステップT2に戻る。

【0036】ステップT6では、入力文字列“S”の“pos”文字目から文字数“keyLen”の文字列を検索キーとして、辞書検索を行う。

【0037】ステップT7では、検索キーが辞書に登録されていたのか否かをチェックし、登録されていればステップT8に進み、登録されていなければステップT9に進む。

【0038】ステップT8では、検索キーが辞書に登録されていたので、得られた情報をトークン・リストに登

10

20

30

40

50

録し、“pos”の値を今検索したキーの文字数分ほど進めて、ステップT2に戻る。

【0039】ステップT9では、検索キーが辞書に登録されていなかったため、キーの文字数を1文字減らし、ステップT4に戻って再度辞書検索を行う。

【0040】ステップT2以降の処理を繰り返し、入力文(a)の最後の文字まで辞書検索を終えた状態のトークン・リストを図6に示す。図6において、P1およびP2は1文字で分割されているが、双方ともステップT5で切り出された未知語ではなく、ステップT8で切り出されたものである。仮に、「上」という文字が辞書に登録されていなかったとすると、P1の品詞は未定で、未知語としてステップT5で切り出される。この2つの違いは、後述する未知語処理で、前者はローマ字表記の読みが出力できるが、後者は読みが出力できないということにある。

【0041】入力文(a)に対しての辞書検索が終了すると、図4に戻り、ステップS4で、ステップS3における文の分割処理が正常に行えたのか否かをチェックする。正常であればステップS6に進み、異常があれば、文分割処理でエラーが生じた旨を通知し(ステップS5)、処理を終える。なお、ステップS5におけるエラーの通知は、表示装置22にその旨のメッセージを表示させることなどが考えられる。

【0042】ステップS6では、トークン認定部12が未知語処理部11を用いて、文分割処理の結果(図6に示すトークン・リスト)を入力として未知語処理を行う。この処理の流れを図7に示すフローチャートを用いて説明する。

【0043】図7に示すように、まず、ステップU1では、変数“P”にトークン・リストの先頭ポインタをセットする。

【0044】ステップU2では、“P”が指すトークンが1文字であるのか否かをチェックし、1文字であればステップU3に進み、2文字以上であればステップU4に進む。

【0045】ステップU3では、“P”が漢字、平仮名、カタカナ、英数字のうちのいずれかであるのか否かをチェックし、いずれかであればステップU6に進み、いずれでもなければステップU4に進む。

【0046】ステップU4では、“P”に次のトークン・データが存在するの否かをチェックし、存在すれば、ステップU5において“P”に次のトークン・データのポインタをセットし、ステップU2に戻る。“P”に次のトークン・データが存在しなければ、処理を終える。

【0047】ステップU6からステップU15までは、一文字のみで切り出された同一字種の連続を検出する。この処理によって、変数“P”は、同一字種の連続の先頭のトークン・データを指し、変数“L”は最後のト

ケン・データを指している。

【0048】ステップU6では、“P”に次のトークン・データが存在するの否かを再度チェックし、存在しなければ“L”に“P”をセットし(ステップU11)、ステップU13に進む。“P”に次のトークン・データが存在すればステップU7に進む。

【0049】ステップU7では、“L”に“P”の次のトークン・データのポインタをセットする。

【0050】ステップU8では、“L”が指すトークンが1文字で、且つ、“L”が指すトークンと“P”が指すトークンが同一の字種か否かをチェックし、同一の字種であればステップU9に進み、同一の字種でなければステップU12に進み、“L”に“L”の前のトークン・データのポインタをセットし、ステップU13へ進む。

【0051】ステップU9では、“L”に次のトークン・データが存在するの否かをチェックし、存在すれば“L”に次のトークン・データのポインタをセットし(ステップU10)、ステップU8へ戻る。存在しなければステップU13に進む。

【0052】ステップU13では、“P”から“L”までのトークンを一つのトークンとして連結し、ローマ字読みを生成する。この“P”から“L”までのトークンの字種が漢字であれば、辞書群2のローマ字読みを参照し、トークンの字種が“平仮名”であれば、平仮名・ローマ字変換テーブル3を参照し、トークンの字種がカタカナであれば、カタカナ・ローマ字変換テーブル14を参照して、それぞれローマ字読みを生成することになる。

【0053】また、トークンの字種が英数字であれば、変換の必要はなく、そのままが良い。なお、固有名詞として処理することを考慮すると、ローマ字読みの先頭文字のみは、大文字にしておくことが望ましい。

【0054】ここで、未知語処理を実行した後のトークン・リストを図8に示す。図6と比較すると、P1およびP2が1つのトークンとしてまとめられていることが分かる。

【0055】図7に示す未知語処理が終了すると、図4に戻り、ステップS7で、未知語処理が正常に行えたのか否かをチェックし、正常であれば、ステップS9で解析結果を出力し、異常が発生したのであれば、ステップS8で未知語処理でエラーが生じた旨を通知し、処理を終える。

【0056】以上に説明したような方法で、日本語文の形態素解析を行うことにより、例えば、「上沼」という文字列が辞書に未登録であったため、「上」と「沼」に分割された場合でも、漢字1文字毎に分割されるのを防ぎ、解析精度を向上させることが可能となる。

【0057】また、ローマ字読みの情報を対訳として用いることにより、日本語から英語に翻訳する機械翻訳に

10

20

30

40

50

において、自然な訳出を行うことが可能となる。ただし、読みは、「jounuma」「kaminuma」「uenuma」など幾つか考えられるが、どの読みを採用するかは、固有名詞として、どの読みがもっともよく使用されるかなどの統計情報を用いて最適な読みを決定する。

【0058】また、「再分割」という文字列が「再」と「分割」に分割された場合なども、漢字トークン認定部7での処理対象として、処理することも可能である。ただし、この場合は未知語ではなく、「再」は、接頭語であるという情報から、「再分割」は固有名詞ではなく、例えば、「分割」に「再」（英語のre-に相当する）の要素が加わったものというような出力にしてもよい。

【0059】

【発明の効果】以上のように本発明によれば、入力される日本語文字列を記憶する入力文記憶手段と、形態素解析に必要な語彙情報を記憶した辞書を用い、日本語文字列をキーとして辞書を検索する辞書検索手段と、辞書検索手段を用いて、入力文記憶手段に記憶された入力文を形態素に分割する文分割手段と、文分割手段により分割された形態素に対し、1文字で形態素となるものが存在していれば、当該形態素と他の形態素とを連結して単語が形成されるのか否かを判断する未知語処理手段と、未知語処理手段の結果に基づいて分割された個々の単位を形態素として認定する形態素認定手段と、を有する構成としたことにより、入力文を分割した際に1文字毎に分割されるのを防ぐことができ、形態素解析の精度を向上させることができる。

【図面の簡単な説明】

【図1】本発明の一実施の形態における日本語形態素解析装置の機能ブロック図

【図2】本発明の一実施の形態における日本語形態素解析装置の回路ブロック図

【図3】本発明の一実施の形態におけるローマ字変換テーブルの説明図

【図4】本発明の一実施の形態における日本語形態素解析装置の動作フローチャート

【図5】本発明の一実施の形態における日本語形態素解析装置の動作フローチャート

【図6】本発明の一実施の形態における文分割処理の結果を示す図

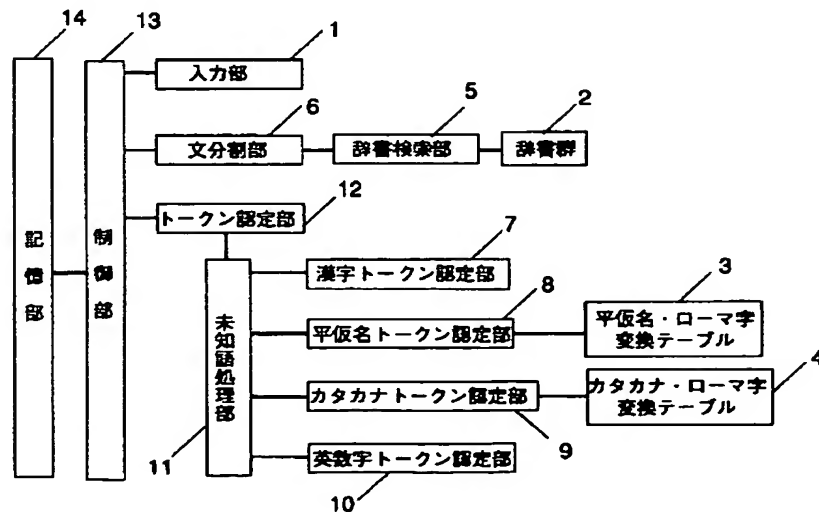
【図7】本発明の一実施の形態における日本語形態素解析装置の動作フローチャート

【図8】本発明の一実施の形態における未知語処理の結果を示す図

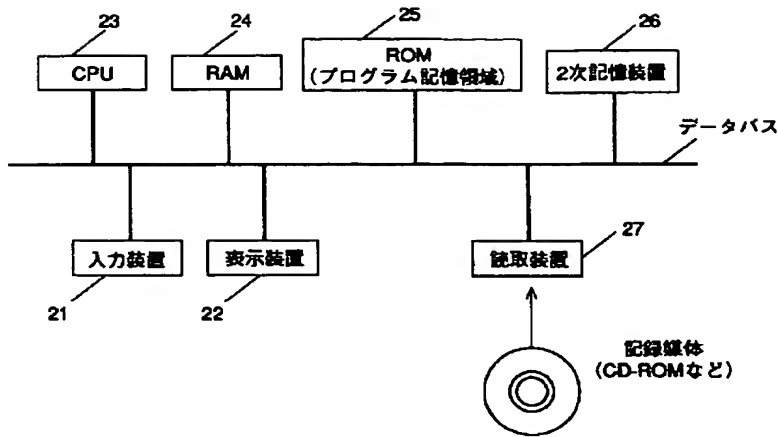
【符号の説明】

- 1 入力部
- 2 辞書群
- 3 平仮名・ローマ字変換テーブル
- 4 カタカナ・ローマ字変換テーブル
- 5 辞書検索部
- 6 文分割部
- 7 漢字トークン認定部
- 8 平仮名トークン認定部
- 9 カタカナトークン認定部
- 10 英数字トークン認定部
- 11 未知語処理部
- 12 トークン認定部
- 13 制御部
- 14 記憶部
- 21 入力装置
- 22 表示装置
- 23 CPU（中央処理装置）
- 24 RAM（ランダム・アクセス・メモリ）
- 25 ROM（リード・オンリー・メモリ）

【図1】



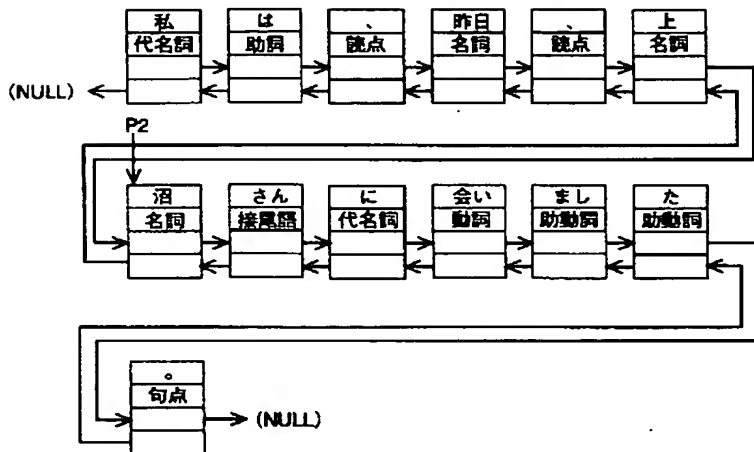
【図 2】



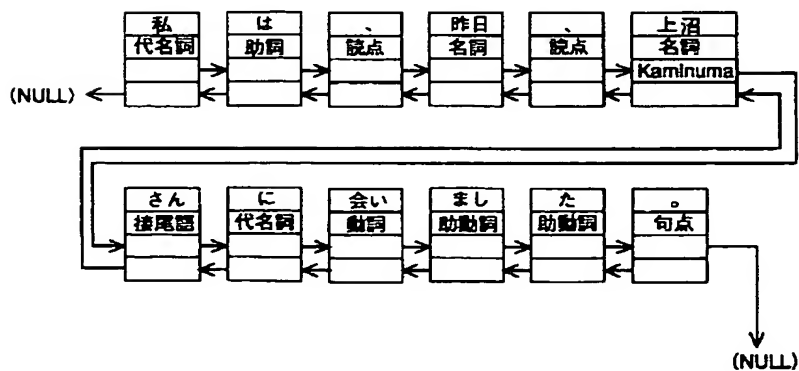
【図 3】

平仮名	カタカナ	ローマ字
あ	ア	a
い	イ	i
う	ウ	u
え	エ	e
お	オ	o
・	・	・
・	・	・
・	・	・
・	・	・
・	・	・
さ	サ	sa
し	シ	shi
・	・	・
・	・	・
ば	バ	ba
び	ビ	bi
ぶ	ブ	bu
・	・	・
・	・	・
・	・	・
・	・	・
・	・	・

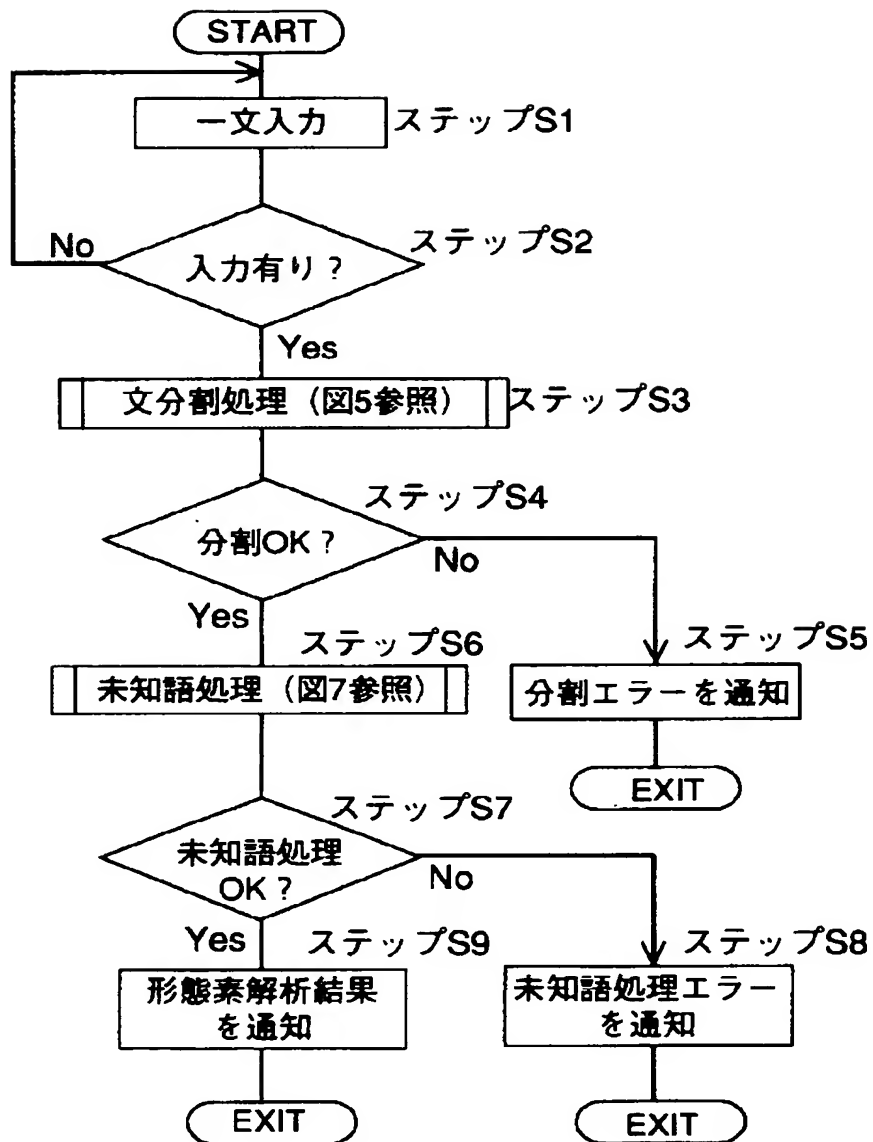
【図 6】



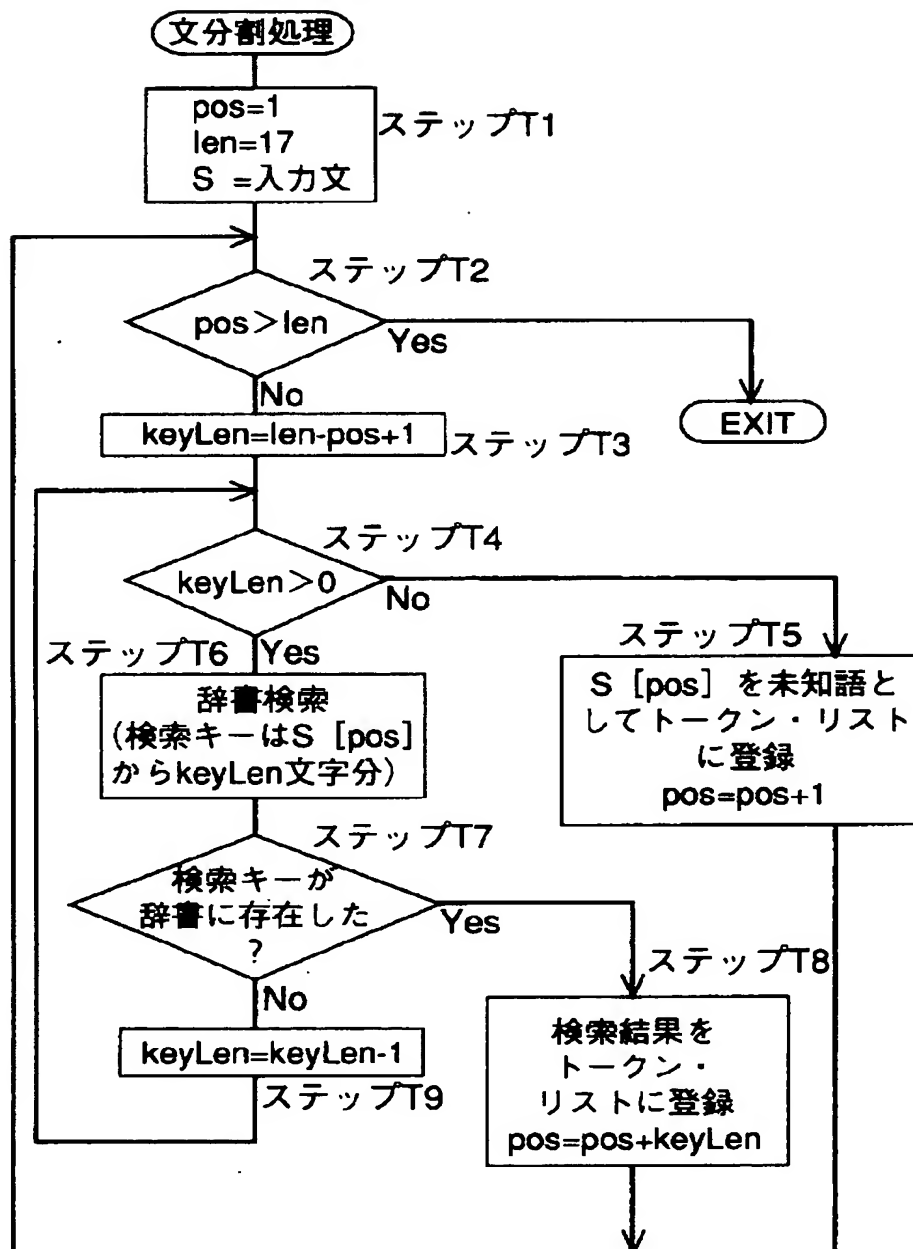
【図 8】



【図4】



【図5】



【図7】

